OCTOGON Mathematical Magazine
Vol. 33. No. 2B. October, 2025.

OCTOGON MATHEMATICAL MAGAZINE
Vol. 33, No.2, October 2025, pp 1312-1333
Print: ISSN 1222-5657, Online: ISSN 2248-1893
https://web.uni-miskolc.hu/∼matsefi/Octogon/

Proposing a Neutrosophic Fuzzy Logic-Based Framework for Dataset Construction in Emotional Intelligence Applications

Rmm Pradeep and Florentin Smarandache [28]

Dedicated to Themistocles M. Rassias on the occasion of his 75th Birthday

ABSTRACT. Human emotional responses are inherently uncertain, ambiguous, and subjective, making it challenging for traditional binary and fuzzy logic approaches to accurately capture this complexity in dataset construction for emotional intelligence (EI) applications.

This limitation impedes the development of EI systems that effectively interpret and respond to nuanced human emotions. To address this challenge, this paper aims to propose a neutrosophic fuzzy logic-based framework that explicitly models and preserves human uncertainty during dataset creation.

The methodology embeds neutrosophic logic into the early stages of EI modelling by designing uncertainty-aware questionnaires that obtain responses as neutrosophic triplets; truth, indeterminacy, and falsity. These responses are normalized, verified for consistency, and converted to scalar scores using entropy-based weighting.

A case study on modelling "Range Anxiety" in electric vehicle purchase decisions illustrates the framework's practical application, including outlier handling and weight calculation.

Results demonstrate that the approach successfully retains individual perceptions and uncertainties without discarding contradictory data. Entropy analysis reveals that falsity plays a significant role in differentiating emotional states,

[28]Received: 23.08.2025
2010 *Mathematics Subject Classification.* 03B15.
Key words and phrases. Neutrosophic Fuzzy Logic, Human Uncertainty, Emotional Intelligence, Dataset Creation, Subjective Data Modelling .

while the resulting scalar scores provide richer features than conventional datasets.

In conclusion, the proposed framework offers a replicable, transparent method for creating uncertainty-aware EI datasets, supporting more emotionally intelligent AI systems. Future work includes validating the framework on diverse real-world datasets to confirm its robustness and applicability.

I INTRODUCTION

A. Background

Today, capturing human behaviours and emotions is getting vital due to research agendas focusing on the Emotional Intelligence of Artificial Intelligence knowledge area (Salovey and Mayer, 1990; Picard, 1997; König and Mehrotra, 2025). However, human behaviour and emotions are based on dynamic conditions, so representing them as data is getting more difficult (Russell and Barrett, 1999; Ekman and Cordaro, 2011). Especially when it comes to developing datasets to be trained for modelling in artificial intelligence applications, there is a demanding requirement to reflect real world complexities and this urges us to develop models that can represent vague, subjective, or indeterminate data.

When considering these complexities in modelling, traditional binary representation of status will not be enough. For example, when one person considers a 1.5 meter person as tall, another may consider it as not tall. So to handle such perspectives, fuzzy logic and its extended versions are available as well as providing substantial assistance (Zadeh, 1965).

However, due to the limitations of crisp fuzzy logic, researchers have developed more advanced fuzzy logic models that can handle more complex vague situations.

There, the neutrosophic logic, the latest development in fuzzy logic that represents truth (T), indeterminacy (I), and falsity (F) independently and simultaneously, is getting attention in this work (Smarandache, 2018).

B. The Problem

Nevertheless, most textbooks and theoretical guides start with a fixed fuzzy level or tri-value already captured from a real-world dataset. The question of

how to create a dataset with such fuzzy uncertainty is often skipped in the books and guides.

This is leading to a gap between theory and practical data engineering. Therefore, when implementing fuzzy theory to handle such situations, computing students and researchers face practical problems in their attempts to practice

C. Aim and Scope

The present work aims to fill the gap between theoretical modelling and practical data collection by introducing a structured method for computing students and researchers to construct datasets that include fuzzy uncertainty from the outset.

The proposed framework outlines how to design uncertainty-aware questionnaires, encode human subjectivity, and derive neutrosophic values in a traceable and replicable manner.

The scope of this study focuses on preserving individual perceptions and uncertainties within data, rather than following other common priorities that aim to reduce uncertainty.

While many approaches seek to eliminate ambiguity, this work recognizes the value of maintaining inherent uncertainty—particularly in domains such as emotional state recognition, behavioural modelling, and human–AI interaction, where traditional crisp datasets often fall short.

By emphasizing the natural variability and subjectivity of human data, the study aims to more accurately capture real-world complexities.

II Theoretical Background and Related Work

A. Dataset Development

When considering the problem and aim of the present work, it is essential first to understand the existing methods for dataset creation in AI modelling alongside how fuzzy logic theory describes and manages uncertainty in data.

As demonstrated in Figure 1, the process of developing datasets for AI

modelling typically begins with defining clear objectives and understanding the problem domain, which sets the foundation for all subsequent activities.

Once the problem is specified, the next critical step is data collection, where relevant data is gathered from diverse sources such as internal databases, public repositories, web scraping, crowdsourcing platforms, and synthetic data generation, especially in privacy-sensitive domains.

This data collection and preparation phase is crucial to ensure high data quality and typically consumes 50 to 80 percent of the project timeline, reflecting the significant investment required to build robust AI systems.(Rahman, 2023; Aggoune and Mihoubi, 2024).

The collected data often includes both structured data (e.g., tabular records) and unstructured data (e.g., text, images, audio), requiring careful consideration of the data types needed for the AI task. Following collection, data preparation and preprocessing are essential to ensure data quality.

This involves cleaning to remove errors and inconsistencies, handling missing values, normalizing formats, and performing feature engineering to extract meaningful variables.

Annotation and labelling of data, often supported by human-in-the-loop systems or AI-assisted tools, are crucial for supervised learning tasks, enabling the model to learn from accurately tagged examples (Kordon, 2020; Gupta, Sehgal and Acken, 2025).

After that, a suitable model selection and development is conducted based on not only the problem type and data characteristics, but also computational resources.

However, the common practice is to employ pre-trained models and transfer learning techniques for quicker application development and to achieve higher accuracy.

Then the selected model is trained on the prepared dataset, with continuous monitoring and tuning of the parameters to optimize the results (Oneto et

al., 2015).

At the last step, the model is tested and validated. These steps are repeated until the model results in substantial output. In this process, ethical considerations also need to be considered when dealing with sensitive emotional or biometric data.

Define Problem Domain and Objectives
Establish clear goals for the AI task
Data Collection
Gather data from diverse sources
Data Annotation and Labeling
Tag data with emotional states and metadata
Ethical and Privacy Considerations
Ensure compliance with regulations
Data Preprocessing and Structuring
Organize data into standardized formats
Train and Fine-Tune AI Models
Use data to improve model capabilities

Figure 1: Pipeline of AI Model Creation
Source : Author, Created with Napkin.ai

B. Human Uncertainty in AI and Emotional Intelligence Applications

Uncertainty and variability are inherent in human decision making, as human decisions extend beyond mere survival instincts and are influenced by

different emotional states. Human feelings such as love, anger, and sadness are particularly difficult to detect because people can conceal their emotions and behave differently than they feel.

Consequently, this behaviour becomes less reliably detectable, making it challenging to capture such data elements for use in datasets (Bland and Schaefer, 2012; Landowska, 2019) Although researchers have attempted to address this by combining multiple modalities such as facial expressions, voice, text, and physiological signals, these complex combinations introduce additional complexity in data fusion and real-time processing (Sanku and Sandhya, 2024).

Not only capturing the data, but also due to ethical concerns being important with human subjects, the data collection and processing is getting more difficult (Kaplan et al., 2023).

Furthermore, there is no universal behaviour of humans, as they are dynamic with individuals, cultures, and situations.

Then, the replication of developed models to another location is getting more challenging.

In addition, the models which are developed to capture human uncertainties have challenges such as poor real-time performance, inability to understand rare emotions, and misinterpretation. Not only that, but also when it attempts to capture very sensitive data, humans tend to deceive the data collectors (Wang et al., 2023).

Therefore, to assist in developing human-centred AI and EI applications, logical foundations have been developed.

C. Logical Foundations of Uncertainty Representation

Data is always a status of an individual based on a logical concept, such as a car is a vehicle but not an animal, but a van is not a car but is a vehicle.

Therefore, each data is having a binary concept, whether it has to be a member or not. So this relates to classical binary logic where the statuses

are described.

Crisp Sets. Crisp set theory is based on classical binary logic, which provides a " member or not" condition as described earlier. Here, membership is considered as absolute - either zero or one with no intermediate values. For example, when recording the skin colour of a woman, it records either " white or not," whereas reality is different.

This approach was too straightforward and insufficiently characterized the gradual transitions or partial membership.

Fuzzy Sets. As an answer to this limitation, fuzzy set theory was introduced by Zadeh (1965). Here, white skin can be recorded not as " white or not," but as a portion where the data can be used to compare the whiteness of skin among the data population.

This conceptualization facilitates modelling vagueness to an extent. However this shows the positivity towards the membership acceptance.

Intuitionistic Fuzzy Sets. Then Atanassov (1986) further developed fuzzy logic, where it captures positivity, negativity and intermediacy.

These parameters were named as membership, non-membership, and hesitation degree. Then intuitionistic fuzzy sets capture the degree of skin colour whiteness, the degree that skin colour is not white, and specifically the degree of confusion about whether the skin colour is white or not.

When assessing one woman's skin colour, an intuitionistic fuzzy set may assign a membership degree of 0.6 where the meaning is skin is more towards white, a non-membership degree of 0.3 where it's less towards non-white skin, and a hesitation degree of 0.1 where there's very minimal hesitation in deciding if the skin is white or not. So this model could be able to model the observation more accurately.

Nevertheless, here the total of the degrees are equal to one, which denotes that human thinking has fully limited capacity. In other words, the limitation of capacity affects all the feelings where if one of the memberships

increases, others are getting affected.

Neutrosophic Sets: However, it is considered that all feelings have their own capacity to increase or decrease without affecting others. Then Smarandache (1999) introduced neutrosophic set theory, which captures truth (T), indeterminacy (I), and falsity (F). These range from 0 to 1 without the constraint that their sum equals 1.

This independence allows neutrosophic sets to model not only uncertainty but also inconsistency and contradiction within data where human beliefs naturally exist.

The simplest example of this type of situation is a woman who loves her husband as well as hates him with the same weight, and there is an unknown feeling regarding him. But no individual feeling is affecting the others. Then due to this more realistic nature, neutrosophic models can be employed in emotional intelligence data creation as they help in modelling contradictory hypotheses, where indeterminacy plays a critical role.

III Proposed Methodology

This study presents a novel methodology that integrates neutrosophic logic into the early stages of AI model development to better capture uncertainty in human-centered data.

By encoding each data point as a neutrosophic triplet—truth, indeterminacy, and falsity—the approach enriches dataset features, enabling machine learning models to handle ambiguity and improve interpretability and robustness. Figure 1 shows the overall AI model development pipeline, from problem definition through evaluation.

This research focuses on embedding neutrosophic logic into data collection and preprocessing without altering the standard workflow.

Encoding data as neutrosophic triplets preserves subjective uncertainty alongside conventional numeric processing.

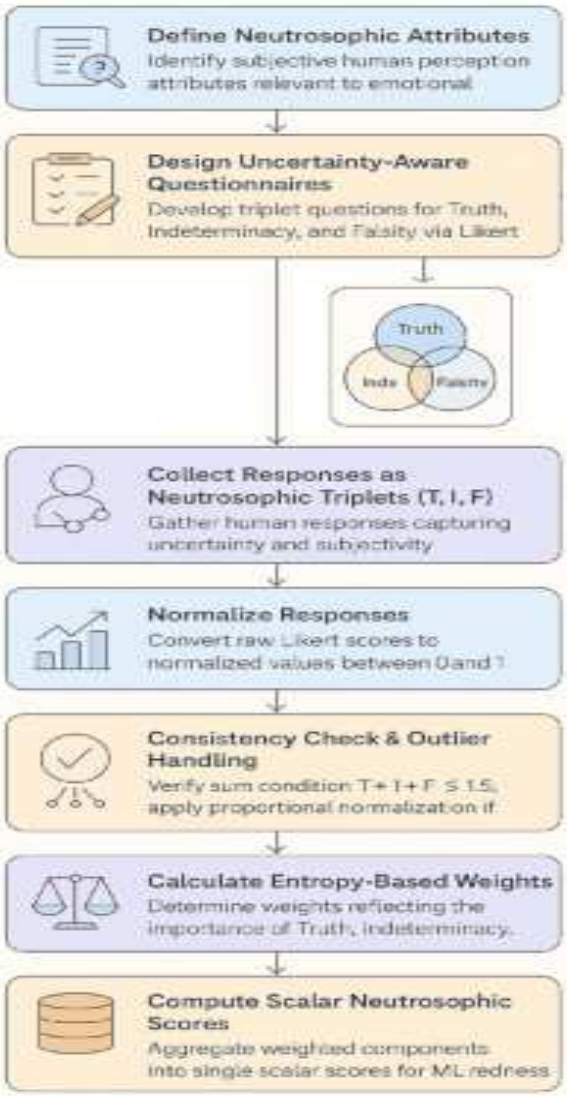
Define Neutrosophic Attributes
Identify subjective human perception attributes relevant to emotional
Design Uncertainty-Aware Questionnaires
Develop triplet questions for Truth, Indeterminacy, and Falsity via Likert
Truth
Collect Responses as Neutrosophic Triplets (T, I, F)
Gather human responses capturing uncertainty and subjectivity
Normalize Responses
Convert raw Likert scores to normalized values between 0 and 1
Consistency Check & Outlier Handling
Verify sum condition T+I+F ≤ 1.5, apply proportional normalization if
Calculate Entropy-Based Weights
Determine weights reflecting the importance of Truth, Indeterminacy,
Compute Scalar Neutrosophic Scores
Aggregate weighted components into single scalar scores for ML readiness

Figure 2: Proposed Methodology
Source : Author, Created with ChapGPT.ai

To illustrate the proposed process, Figure 2 provides a detailed flow diagram of the methodology, outlining steps including defining neutrosophic attributes, designing uncertainty-aware questionnaires, collecting and normalizing responses, consistency checking, entropy-based weighting, and finally computing scalar neutrosophic scores to structure the enriched

dataset. This systematic approach captures individual perceptions and uncertainty, supplying machine learning algorithms with richer, uncertainty-aware features for improved modelling.

1). Defining Neutrosophic Attributes.

First, it is necessary to identify the human perception attributes within the objective function of the model. Then, it must determine how to represent these attributes as neutrosophic triplets. This process requires careful selection of attributes and clear communication with respondents to ensure they understand and can effectively differentiate the components. Lets consider one such attribute is x, and is represented as a neutrosophic triplet (Equation 1):

$$x = (T, I, F) \tag{1}$$

where

T : degree to which the attribute is believed true,
I : degree of uncertainty or hesitation,
F : degree to which the attribute is believed false.
These components are independent, and their sum can range between 0 and 3.

2). Developing Questionnaire Items for Each Attribute

To capture the three components of the neutrosophic triplet, three distinct questions (rather statement) are developed per attribute.

a. Truth (T): To evaluate the level of positivity
b. Indeterminacy (I) : To evaluate the level of uncertainty of emotion on the object
c. Falsity (F): To evaluate the level of negativity
Then setup the answer for the question/statement on a Likert scale (e.g., 1 to 5). However the clear instructions and pilot testing are required to help respondents distinguish indeterminacy from truth and falsity.

3). Collecting Responses as Neutrosophic Triplets

Each respondent r provides answers to the three questions per attribute,

forming a neutrosophic triplet (Equation 2)

$$x^{(r)} = \left(T^{(r)}, I^{(r)}, F^{(r)}\right) \tag{2}$$

where

T(r): degree to which respondent r believes the attribute is true
I(r): degree of uncertainty or hesitation expressed by respondent r
F(r): degree to which respondent r believes the attribute is false
This preserves individual perceptions and uncertainties, which are critical for personalized machine learning modelling. However, responses may contain noise or inconsistencies, requiring preprocessing.

4). Normalizing Responses

Then it needs raw Likert-scale responses $s \in \{1, 2, ..., m\}$ are to be normalized to the unit interval [1] to ensure comparability and prepare for mathematical operations (Equation 3):

$$v = \frac{s-1}{m-1} \tag{3}$$

where

s :raw Likert-scale response given by a participant (an integer from 1 to mmm)
m: total number of points on the Likert scale (e.g., 5 for a 5-point scale)
v: normalized score in the interval [0,1][0, 1][0,1]
Normalization is applied independently to each component T, I, and F.

5) Consistency Check and Adjustment

Since T,I,F are independent, their sum can exceed 1 or even 3, which may cause interpretability issues or illogical values. To maintain a valid neutrosophic representation, the sum is checked (Equation 4)

$$T + I + F \leq 1.5 \tag{4}$$

If this is violated, proportional normalization is applied (Equation 5).

$$T = \frac{T}{T+I+F}, I = \frac{I}{T+I+F}, F = \frac{F}{T+I+F} \tag{5}$$

This scaling preserves the relative proportions while bounding the sum. This step is optional and should be applied only if necessary, after data inspection, as it may slightly distort original responses.

6). Computing Scalar Neutrosophic Scores

Most machine learning algorithms require scalar numerical features. Therefore, each neutrosophic triplet is converted into a scalar score ($S^{(r)}$) per respondent r and attribute using a weighted aggregation (Equation 6).

$$S^{(r)} = aT^{(r)} + \beta\left(1 - I^{(r)}\right) + \gamma\left(1 - F^{(r)}\right) \tag{6}$$

where

S(r): scalar neutrosophic score representing the overall degree of belief adjusted for uncertainty and falsity for respondent r
T(r): degree to which respondent r believes the attribute is true
I(r): degree of uncertainty or hesitation expressed by respondent r
F(r): degree to which respondent r believes the attribute is false
$\alpha, \beta, \gamma \geq 0$ and $\alpha + \beta + \gamma = 1$: weights controlling the influence of each component

7). Structuring the Dataset for Machine Learning:

The last step of the proposal is to construct the dataset with the scalar scores, S(r), for each respondent and attribute.

IV. Case Study: Modelling Range Anxiety in EV Purchase Decisions

Let's consider the application of the proposed neutrosophic methodology to a real-world case: predicting the likelihood of a customer purchasing an electric vehicle (EV) based on the subjective attribute Range Anxiety. Range Anxiety reflects the fear or concern about the limited driving range of an EV and its influence on purchase decisions. This attribute is inherently subjective and varies across individuals, making it suitable for neutrosophic modelling that captures belief, hesitation, and disbelief.

1). Step 1: Defining Neutrosophic Attribute

It focuses on the attribute Range Anxiety, represented as a neutrosophic

triplet as per the Equation 1, where:

• T: degree of belief in low range anxiety,
• I: degree of uncertainty or hesitation,
• F: degree of belief in high range anxiety.

2). Step 2: Developing Questionnaire Items

To capture the three components, three statements were developed, each rated on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree):
• Truth (T): " I have confidence that the range of the EV can be managed by myself and is sufficient for my needs."
• Indeterminacy (I): " I wonder whether the range of the EV influences my buying decision or not."
• Falsity (F): " The range is one of the nightmares when I decide to use my vehicle."

3). Step 3: Collecting Responses

Imagin, six respondents provided their perceptions as Likert scores as shown in Table 1:

Table 1: Initial data captured through questioner

Respondent	*T Response*	*I Response*	*F Response*
1	4	2	2
2	2	3	4
3	3	4	2
4	5	4	3
5	2	2	1
6	5	5	5

4). Step 4: Normalizing Responses

Each raw Likert response S was normalized to 1 using Equation 3. Accordingly resulting normalized values are shown in Table 2.

Table 2: Normalised Responses

Respondent	T	I	F
1	0.75	0.25	0.25
2	0.25	0.50	0.75
3	0.50	0.75	0.25
4	1.00	0.75	0.50
5	0.25	0.25	0.00
6	1.00	1.00	1.00

5). Step 5: Consistency Check and Suitability Assessment

The sum S=T+I+F was computed for each respondent to check validity and assess suitability based on the condition $S \leq 1.5$ and results are shown in Table 3.

Table 3: Suitability Assessed Responses

Res.	T	I	F	$Sum\,S$	$Suitability*$
1	0.75	0.25	0.25	1.25	*Suitable*
2	0.25	0.50	0.75	1.50	*Suitable*
3	0.50	0.75	0.25	1.50	*Suitable*
4	1.00	0.75	0.50	2.25	*Outlier*
5	0.25	0.25	0.00	0.50	*Suitable*
6	1.00	1.00	1.00	3.00	*Outlier*

Suitability*:

Suitable: Respondents whose neutrosophic triplets satisfy the sum condition and can be used directly in modelling.

Outlier: Respondents violate the sum condition, indicating inconsistent or contradictory responses.

Then Outliers' triplets were scaled proportionally to satisfy $S \leq 1.5$ using Equation 5 and the normalised values are shown in Table 4.

Table 4: Normalised outliers

Res.	$Original(T, I, F)$	$Sum\,S$	$Normalized(T, I, F)$
4	$(1.00, 0.75, 0.50)$	2.25	$(0.44, 0.33, 0.22)$
6	$(1.00, 1.00, 1.00)$	3.00	$(0.33, 0.33, 0.33)$

6). Step 6: Computing Scalar Neutrosophic Scores

In this step, Equation 6 is utilized to compute the scalar neutrosophic scores, which requires calculating the weights α, β and γ corresponding to the Truth, Indeterminacy, and Falsity components, respectively.
For this demonstration, the Entropy Weight Method is employed to objectively determine these weights based on the variability of the data. The entropy calculation uses the dataset shown in Table 4; however, outliers (Respondents 4 and 6) are excluded from the entropy computation to avoid bias. Although these outliers have been proportionally normalized (as shown in Table 4) to satisfy the neutrosophic sum condition, this normalization only rescales their values without resolving the underlying inconsistencies or contradictions in their responses. Including these outliers in entropy calculation could distort the weight determination by artificially inflating or deflating the variability of the neutrosophic attributes.
Therefore, to maintain the robustness and reliability of the entropy-based weighting, only suitable respondents (as shown in Table 5), whose neutrosophic triplets satisfy the sum condition without contradiction, are included in the entropy calculation. This ensures that the weights reflect the true information content and natural variability of the consistent data.

Table 5: Neutrosophic Data Used for Entropy Calculation

Respondent	T	I	F
1	0.75	0.25	0.25
2	0.25	0.50	0.75
3	0.50	0.75	0.25
5	0.25	0.25	0.00

Then, to begin, the neutrosophic attribute values x_{ij} for each respondent i and attribute $j \in \{T, I, F\}$ are normalized into probability distributions as shown in Equation 7.

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} \tag{7}$$

where
p_{ij}: normalized value (interpreted as a probability) of attribute jj for respondent ii,

x_{ij}: original normalized neutrosophic value, n: number of respondents considered (excluding outliers).
This normalization ensures that for each attribute j, the sum of pij over all respondents equals 1 (Table 6).

Table 6: Normalized Probabilities for the Suitable Respondents

Respondent	p_T	p_I	P_F
1	0.429	0.143	0.200
2	0.143	0.286	0.600
3	0.286	0.429	0.200
5	0.143	0.143	0.000
$\sum$	1.001	1.001	1.000

Next, the product $p_{ij} \ln(p_{ij})$ is calculated for each respondent and attribute, treating zero probabilities as zero (since $0 \times \ln(0)=0$). The natural logarithm values are used here, and the results are summarized in Table 7.

Table 7: Computation of $p_{ij} \ln(p_{ij})$

Respondent			
1	-0.3629	-0.2783	-0.3219
2	-0.2783	-0.3577	-0.3065
3	-0.3577	-0.3629	-0.3219
5	-0.2783	-0.2783	0
Sum	-1.2772	-1.2772	-0.9503

After that, the normalization constant k is calculated to scale the entropy values between 0 and 1.
Since there are 4 respondents,

$$k = \frac{1}{1,3863} \approx 0.7123$$

Then, the entropy E_j for each attribute is computed using the formula shown in Equation 8 and results are shown at Table 8.

$$E_j = -k \sum_{i=1}^{n} p_{ij} \ln(p_{ij}) \tag{8}$$

where

E_j: entropy of attribute jj,
$k = \frac{1}{???(n)}$: normalization constant that ensures E_j lies between 0 and 1,
ln: the natural logarithm, is defined to be 0 if p_{ij}= 0 (by continuity).

Table 8: The calculated entropy values

Attribute	Calculation	Entropy E_j
Truth (T)	-0.7213×(-1.2772)	0.921
Indeterminacy (I)	-0.7213×(-1.2772)	0.921
Falsity (F)	-0.7213×(-0.9503)	0.685

Next, the degree of diversification D_j, which measures the contrast or information content of each attribute, is calculated using $D_j = 1 - E_j$ and results are shown in Table 9.

Table 9: The calculated Degree of Diversification

Attribute	Degree of Diversification D_j
Truth (T)	1-0.921=0.079
Indeterminacy (I)	1-0.921=0.079
Falsity (F)	1-0.685=0.315

In next step, the weights w_j for each attribute are calculated by normalizing the diversification degrees using the Equation 9.

$$w_j = \frac{D_j}{\sum_i D_j} \tag{9}$$

The resulting weights are shown in Table 10.

Table 10: The entropy-derived weights

Attribute	Weight w_i	Percentage	Related part in Equation 6
Truth (T)	0.167	16.7%	α
Indeterminacy (I)	0.167	16.7%	β
Falsity (F)	0.666	66.6%	γ

These entropy-derived weights indicate that Falsity is the most informative component in the dataset, followed by Truth and Indeterminacy, which have equal weights.

These weights ($\alpha = 0.17$, $\beta = 0.17$, and $\gamma = 0.66$) are then applied in the scalar neutrosophic score aggregation formula (Equation 6) to balances belief (Truth), uncertainty (Indeterminacy), and disbelief (Falsity) into a single scalar score as shown in Table 11.

Table 11: Scalar Neutrosophic Score Calculation

Res.	T Response	Score
1	0.5x0.75+0.3x(1-0.25)+0.2x(1-0.25)=0.375+0.225+0.15	0.75
2	0.5x0.75+0.3x(1-0.50)+0.2x(1-0.75)=0.125+0.15+0.05	0.33
3	0.5x0.50+0.3x(1-0.75)+0.2x(1-0.25)=0.25+0.075+0.15	0.48
4	0.5x0.44+0.3x(1-0.233)+0.2x(1-0.22)=0.22+0.201+0.156	0.58
5	0.5x0.25+0.3x(1-0.25)+0.2x(1-0.00)=0.125+0.225+0.20	0.55
6	0.5x0.33+0.3x(1-0.33)+0.2x(1-0.33)=0.165+0.201+0.134	0.50

7). Step 7: Structuring the Dataset for Machine Learning

The final dataset consists of the scalar neutrosophic scores computed above (Table 12)

Table 12: Final Range Anxiety Dataset

Res.	RA_Certainty	RA_Indeterminacy	RA_Negation	RA_NeutrosophicScore
1	0.75	0.25	0.25	0.75
2	0.25	0.5	0.75	0.33
3	0.50	0.75	0.25	0.48
4	1.00	0.75	0.50	0.58
5	0.25	0.25	0.00	0.55
6	1.00	1.00	1.00	0.50

V Results and Discussion

A. Neutrosophic Data Construction and Outlier Handling

The application of the proposed neutrosophic framework to the Range Anxiety case study revealed several key insights regarding human uncertainty preservation. The initial collection and normalization of Likert-scale responses generated neutrosophic triplets (T, I, F) for each respondent (Table 3). The consistency check, based on the sum condition

$T + I + F \leq 1.5$, identified respondents 4 and 6 as having highly inconsistent or contradictory responses, categorizing them as "outliers" (Table 4). Rather than removing these outliers, their values were proportionally normalized (Table 5) to fit within the neutrosophic sum constraint. This approach demonstrated that the dataset retained all individual perceptions, including those that conventional outlier removal techniques would typically discard, thus preserving a more complete reflection of real-world human data.

B. Entropy-Based Weight Determination

The Entropy Weight Method successfully determined the relative influence of Truth, Indeterminacy, and Falsity components on the final scalar score. The identified outliers (respondents 4 and 6) were temporarily excluded from the entropy computation (Table 6) to prevent artificial distortion of the derived weights. The calculated entropy values (Table 8) and degrees of diversification (Table 9) for the suitable respondents yielded entropy-derived weights of $\alpha = 0.17$ for Truth, $\beta = 0.17$ for Indeterminacy, and $\gamma = 0.66$ for Falsity (Table 10). These results indicate that the "Falsity" component (representing high range anxiety) was the most informative in differentiating responses within this dataset.

C. Final Scalar Neutrosophic Scores and Dataset Structuring

The computation of scalar neutrosophic scores for all respondents, including the adjusted outliers, using Equation 6 (Table 11) ensured that every data point contributed to the final machine learning-ready dataset. The final structured dataset (Table 12) includes all respondents, preserving their raw neutrosophic components alongside their calculated scalar scores and a flag indicating initial suitability. This comprehensive dataset reflects the full spectrum of individual perceptions and uncertainties, demonstrating that inherent contradictions and hesitations can be acknowledged and integrated rather than eliminated. The method successfully retained all respondents in the eventual analysis, incorporating the full range of human feelings.

VI. Conclusion

The application of the proposed neutrosophic framework to the Range Anxiety case study successfully preserved human uncertainty throughout data construction and preprocessing. Instead of discarding inconsistent responses as outliers, the framework employed proportional normalization to

retain these valuable perceptions, maintaining dataset completeness. The entropy-based weighting highlighted the dominance of the "Falsity" component in differentiating emotional states related to range anxiety, underscoring the method's capacity to capture meaningful subjective ambiguity.

Crucially, final scalar neutrosophic scores were computed for all respondents—including adjusted outliers—resulting in a comprehensive, machine learning-ready dataset that integrates diverse and sometimes contradictory emotional responses. This approach affirms the framework's ability to embrace the complexity of human emotions rather than oversimplifying uncertainty, thereby providing richer features for subsequent modelling and analysis.

Overall, these results demonstrate the effectiveness of the neutrosophic methodology in creating robust, uncertainty-aware datasets that reflect the nuanced realities of human emotional data. Future evaluation across broader datasets and application domains will further validate its practical value for emotional intelligence and related AI modelling tasks.

REFERENCES

[1] Aggoune, A. and Mihoubi, Z. (2024) 'Towards Efficient Dataset Development: A Case Study of M2Q2+ in Movie QA Systems', in 2024 International Conference on Advanced Aspects of Software Engineering (ICAASE), pp. 1–5. Available at https://doi.org/10.1109/ICAASE64542.2024.10850962.

[2] Bland, A.R. and Schaefer, A. (2012) 'Different varieties of uncertainty in human decision-making', Frontiers in Neuroscience, 6(JUN), pp. 1–11. Available at: https://doi.org/10.3389/fnins.2012.00085.

[3] Ekman, P. and Cordaro, D. (2011) 'What is meant by calling emotions basic', Emotion Review, 3(4), pp. 364–370. Available at: https://doi.org/10.1177/1754073911410740.

[4] Gupta, P., Sehgal, N.K. and Acken, J.M. (2025) 'Practical Aspects in Machine Learning', in Introduction to Machine Learning with Security: Theory and Practice Using Python in the Cloud. Cham: Springer International Publishing, pp. 281–330. Available at: https://doi.org/10.1007/978-3-031-59170-9_9.

[5] Kaplan, B.A. et al. (2023) 'Data handling: ethical principles, guidelines, and recommended practices', in D.J. Cox et al. (eds) Research Ethics in Behavior Analysis. Academic Press, pp. 191–214. Available at: https://doi.org/https://doi.org/10.1016/B978-0-323-90969-3.00006-2.

[6] König, P. and Mehrotra, S. (2025) " Emotional Responses to Artificial Intelligence Systems: A Systematic Review", SSRN [Preprint]. Available at: https://doi.org/10.2139/ssrn.5189667.

[7] Kordon, A.K. (2020) 'Data Preparation', in Applying Data Science: How to Create Value with Artificial Intelligence. Cham: Springer International Publishing, pp. 221–249. Available at: https://doi.org/10.1007/978-3-030-36375-8_8.

[8] Landowska, A. (2019) 'Uncertainty in emotion recognition', Journal of Information, Communication and Ethics in Society, 17(3), pp. 273–291. Available at: https://doi.org/10.1108/JICES-03-2019-0034.

[9] Oneto, L. et al. (2015) 'Model selection for Big Data: Algorithmic stability and Bag of Little Bootstraps on GPUs', 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2015 - Proceedings, (April), pp. 261–266.

[10] Picard, R.W. (1997) Affective computing, Affective computing. Cambridge, MA, US: The MIT Press.

[11] Rahman, A. (2023) '9 - Data collection, wrangling, and pre-processing for AI assurance', in F.A. Batarseh and L.J. Freeman (eds) AI Assurance. Academic Press, pp. 321–338. Available at: https://doi.org/https://doi.org/10.1016/B978-0-32-391919-7.00022-6.

[12] Russell, J.A. and Barrett, L.F. (1999) 'Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant.', Journal of personality and social psychology, 76(5), pp. 805–819. Available at: https://doi.org/10.1037//0022-3514.76.5.805.

[13] Salovey, P. and Mayer, J.D. (1990) 'Emotional Intelligence', Imagination, Cognition and Personality, 9(3), pp. 185–211. Available at: https://doi.org/10.2190/DUGG-P24E-52WK-6CDG.

[14] Sanku, S.R. and Sandhya, B. (2024) 'Multi-Modal Emotion Recognition Feature Extraction and Data Fusion Methods Evaluation', International Journal of Innovative Technology and Exploring Engineering, 13(10), pp. 18–27. Available at: https://doi.org/10.35940/ijitee.j9968.13100924.

[15] Smarandache, F. (2018) 'Extension of Soft Set to Hypersoft Set , and then to Plithogenic Hypersoft Set', Neutrosophic Sets and Systems, 22, pp. 168–170.

[16] Wang, J.Z. et al. (2023) 'Unlocking the Emotional World of Visual Media: An Overview of the Science, Research, and Impact of Understanding Emotion: Drawing Insights From Psychology, Engineering, and the Arts, This Article Provides a Comprehensive Overview of the Field of Emotion Analysis in Visual Media and Discusses the Latest Research, Systems,

Challenges, Ethical Implications, and Potential Impact of Artificial Emotional Intelligence on Society.', Proceedings of the IEEE. Institute of Electrical and Electronics Engineers, 111(10), pp. 1236–1286. Available at: https://doi.org/10.1109/JPROC.2023.3273517.
[17] Zadeh, L.A. (1965) 'Fuzzy sets', Information and Control, 8(3), pp. 338–353. Available at: https://doi.org/https://doi.org/10.1016/S0019-9958(65)90241-X.
[18] Octogon Mathematical Magazine (1993-2025)

Deparment of Information Technology,
Faculty Computing,
General Sir John Kotelalwal Defence University,
Ratmalana, Sri Lanka
E-mail: pradeep@kdu.ac.lk

Emeritus Professor,
Mathematics, Physics, and Natural Science Division,
University of New Mexico,
705 Gurley Ave., Gallup, NM 87301, USA